STATS 101

Introduction to Applied Statistical Methods

昆山杜克大学
DUKE KUNSHAN UNIVERSITY

Session 2, 2024-25

Course meeting time: Monday & Wednesday 2:45pm – 5:15pm; Tuesday 2:45pm – 4:00pm
Course meeting location: IB 3106
Course format: Lecture & lab
Academic credit: 4

Instructor's information

Markus Neumann
Assistant Professor of Political Science and Computational Social Science
Email: markus.neumann@dukekunshan.edu.cn
Office: WDR 1104
Office hours: Monday & Wednesday 10am-12pm, or by appointment
Professional website: https://markusneumann.github.io/

I am an Assistant Professor of Political Science and Computational Social Science at DKU. My research revolves around the application of statistics and machine learning methods to social science data, especially text, images and audio. The substantive focus of my research is political advertising.

Getting in touch with me
Feel free to send me an email about any questions you may have. Make sure the subject line contains 'STATS101'. During the week, I will try to respond within 24 hours. If you don't receive a response in that time, feel free to email me again. You can also come to my office hours, or make an appointment if the office hours times don't work for you.

What is this course about?

How can we use data to shed light on age-old and new human problems such as pollution, discrimination, and economic growth? How can we be "sure" that the evidence we have points us in the right direction? How meaningful are our findings? Do our results suggest the relationships we find between factors such smoking and cancer are meaningful or meaningless? How would we know? How should one properly display and explain your statistical results to these important issues?

This class introduces you to the tools and concepts that begin to tackle these questions. We will cover topics such as data summaries, sampling, data analysis, production of graphical displays, and regression techniques. The goal at the end of the course is that you will be able to conduct basic data manipulation, know how to properly summarize and display data, and make basic statistical inferences using real datasets.

The emphasis in the course will not be on learning mathematical formulas related to statistics but rather to develop an intuitive understanding of statistical inference and measures of uncertainty. For those interested in further study of statistics in a more rigorous way, you may also consider taking the following courses:

> Math 205: the mathematical foundations of statistics
> Econ 203: advanced study of modern regression techniques for Economists
> PoliSci 301: Statistical techniques to infer causation
> Social Sciences 320: advanced statistical techniques applied to real-world problems

What background knowledge do I need before taking this course?

Because this is a 100-level course, no prerequisite course or background knowledge is required.

What will I learn in this course?

Upon completing the course, you will develop the following abilities:

- Intuitively interpret statistics in course materials and in the larger world
- Become a statistics results producer in addition to a statistics consumer
- Assess when and how to use statistics to answer specific questions in the social sciences
- Analyze how previously learned problems can be answered with statistical methods
- Apply statistical methods to future social science coursework and capstone project
- Judge how appropriately statistics are used in everyday life when reading the news, business reports, and other real-world applications

In support of this you will be able to:

- Understand and interpret basic statistical properties of data (confidence intervals, t-tests, etc.)
- Identify when various statistical tests are appropriate given a specific dataset
- Formulate testable hypotheses in the data and learn how to execute those tests
- Interpret statistical results to understand both significance of the results and their substantive impact
- Illustrate statistical results with appropriate and clear graphical displays that provide meaning to the reader
- Evaluate critically other, published, statistical work with the skills and techniques learned in class

What will I do in this course?

- Participate and engage in class discussion and activities.
- Follow along with statistical and programming examples.
- Write code yourself in the lab portion of the class.
- Learn statistical programming in R by completing DataCamp tutorials.

How can I prepare for the class sessions to be successful?

We will be using the following tools in this class:

1. Canvas: Canvas will be used as the central hub of the class, and will contain announcements, class content (lectures and files used therein), assignments (including the place where you will submit your final project), and a discussion forum.
2. R and RStudio: The software packages that implement the R statistical programming language are free and very popular tools for conducting statistical investigations. We will be using R throughout this class.
   - Quarto: Quarto is a flexible document formatting standard that allows you to create nice looking documents that easily mix text and R code output. All of the lectures, as well as your final project will be written inside of Quarto documents. Instructions on how to use the Quarto syntax will be provided during the first week of class.
3. DataCamp: DataCamp has a lot of very useful tutorials that will help you learn how to code in R. These tutorials are part of the class assignments (see below). I will provide you with the link needed so sign up for DataCamp.

What required texts will I need?

Intro Stats, 6th Edition by De Veaux, Velleman, and Bock

How will my grade be determined?

Warmup quizzes (20%):
At the start of each class session there will be a series of short questions to start discussion of the class material. I will drop your two lowest quiz grades.

Midterm (35%):
There will be a midterm on November 13. The midterm will consist of two parts: Statistics and coding. The statistics part will cover chapters 1-9 of the textbook. It will be a closed book exam, and you will not have access to your laptop, the textbook, ChatGPT, etc. During the coding part, you will be using your laptop to write R code to complete exercises similar to those in the lab.

Final project (35%):
The final exam will ask you to analyze a dataset using all the strategies learned in this class and write a 2000-3000 word paper on it.

Data Camp labs (9%):
One lab per week (for the first 6 weeks) on the website Data Camp will be assigned to you; these labs are pass/fail and you will receive full credit if you complete each of the labs by the specified due date.

Syllabus quiz (1%):
A short quiz after the first class regarding the course requirements.

Grade scale:
A+= 98% – 100%; A = 93% – 97.9%; A- = 90% – 92.9%; B+ = 87% – 89.9%; B = 83% – 86.9%; B- = 80% – 82.9%; C+ = 77% – 79.9%; C = 73% – 76.9%; C- = 70% – 72.9%; D+ = 67% – 69.9%; D = 63% – 66.9%; D- = 60% – 62.9%; F = 59.9% and below.

For final grades, .05 is rounded up. For example, a 92.94999 is an A-, a 92.95 is an A. Grades are non-negotiable and can only be changed due to an error in calculation or transcription.

Don't trust Canvas' grade calculation – if you want to know what your current grade is, calculate it yourself.

What are the course policies?

Generative AI Guidelines and Policy
The quizzes and the midterm exam are closed-book assignments. You will be allowed to use ChatGPT for the final project. However, that project is not very amenable to ChatGPT help. Your main grade will be for understanding, annotating, and interpreting your output. Producing the output is the very simple minimal requirement. Also, ChatGPT often makes many coding mistakes. If you rely too much on ChatGPT to code, you will not understand when it makes mistakes and why it makes mistakes.
Finally, while permitted, the use of ChatGPT is still governed by DKU's rules:
- Use of these tools is governed by DKU's Academic Integrity Policy, and students must employ this technology consistent with expectations of the instructor, course, or assessment.
- In any situations in which such tools are used (with or without permission) in the process of completing assignments, students are obliged to cite fully any use of generative AI tools in the formulation of their work, including by preserving a record of the use of the tool as original source material.

- Students are encouraged to save all rough drafts and notes for papers, in case any concerns arise.

DKU also has a [licensed version of ChatGPT](#) that you can (but don't have to) use.

Late Penalties

Since the course moves very quickly, if you are submitting work late that means you are falling behind on other material and it may be difficult for you to recover. Therefore, I have a fairly strict lateness policy.

- The final project and the DataCamp labs are due at 11:59:00 pm. Not 11:59:01 or 11:59:31.
- If it is later than 11:59:00 pm, then the assignment will be assessed a 5% lateness penalty
- If it is later than 12:29:00 am, then the assignment will be assessed a 10% lateness penalty
- If it is later than 11:59:00 pm the next day, the assignment will be assessed a 50% lateness penalty
- If it is later than 2 days from the due date, I will no longer accept the assignment

Please be sure to check that your homework is complete and make sure to submit it a few minutes early. You can submit multiple times on Canvas so make sure you have a nearly complete version uploaded even if you want to keep working on it right up to the deadline. I will not be sympathetic to messages that complain of computer problems when you are trying to submit for the first time at 11:58:51 pm.

Rescheduling of assignment due dates will only be permitted for serious medical and personal matters, **and requires advance notice**. Unless stated otherwise, all assignments are due at 11:59pm China time.

Class Attendance

As outlined in the University Bulletin, regular and punctual class attendance is expected. Absences will only be excused for serious medical and personal matters. If you will be absent from a class for a university-sponsored activity, please make arrangements with me – beforehand – regarding any work you might miss. Note that as of Fall 2024, DKU policy states that students missing 2+ weeks of classes may receive an F or be placed on involuntary leave.

If you will need to miss a class due to observance of a religious holiday, let me know during the first week of class specifically which day(s) you will miss.

Discussion Guidelines

Civility is an essential ingredient for academic discourse. All communications for this course should be conducted constructively, civilly, and respectfully. Differences in beliefs, opinions, and approaches are to be expected. Please bring any communications you believe to be in violation of this class policy to the attention of your instructor. Active interaction with peers and your instructor is essential to success in this course, paying particular attention to the following:

- Be respectful of others and their opinions, valuing diversity in backgrounds, abilities, and experiences.
- Challenging the ideas held by others is an integral aspect of critical thinking and the academic process. Please word your responses carefully and recognize that others are expected to challenge your ideas. A positive atmosphere of healthy debate is encouraged.
- Read your online discussion posts carefully before submitting them.

Academic Integrity

As a student, you should abide by the academic honesty standard of the Duke Kunshan University. Its Community Standard states: "Duke Kunshan University is a community of individuals from diverse cultures and backgrounds. We are dedicated to scholarship, leadership, and service and to the principles of honesty, fairness, respect, and accountability. Members of this community commit to reflecting upon and upholding these principles in all academic and non-academic endeavors, and to protecting and promoting a culture of integrity and trust." For all graded work, students should pledge that they have neither given nor received any unacknowledged aid.

Academic Policy & Procedures

You are responsible for knowing and adhering to academic policy and procedures as published in University Bulletin and Student Handbook. Please note, an incident of behavioral infraction or academic dishonesty (cheating on a test, plagiarizing, etc.) will result in immediate action from us, in consultation with university administration (e.g., Dean of Undergraduate Studies, Student Conduct, Academic Advising). Please visit the Undergraduate Studies website for additional guidance related to academic policy and procedures. Academic integrity is everyone's responsibility.

Academic Disruptive Behavior and Community Standard

Please avoid all forms of disruptive behavior, including but not limited to: verbal or physical threats, repeated obscenities, unreasonable interference with class discussion, making/receiving personal phone calls, text messages or pages during class, excessive tardiness, leaving and entering class frequently without notice of illness or other extenuating circumstances, and persisting in disruptive personal conversations with other class members. Please turn off phones, pagers, etc. during class unless instructed otherwise. Laptop computers may be used for class activities allowed by the instructor during synchronous sessions. If you choose not to adhere to these standards, I will take action in consultation with university administration (e.g., Dean of Undergraduate Studies, Student Conduct, Academic Advising).

Academic Accommodations

If you need to request accommodation for a disability, you need a signed accommodation plan from Campus Health Services, and you need to provide a copy of that plan to me. Visit the Office of Student Affairs website for additional information and instruction related to accommodations.

What campus resources can help me during this course?

Academic Advising and Student Support

Please consult with the instructor about appropriate course preparation and readiness strategies, as needed. Consult your academic advisors on course performance (i.e., poor grades) and academic decisions (e.g., course changes, incompletes, withdrawals) to ensure you stay on track with degree and graduation requirements. In addition to advisors, staff in the Academic Resource Center can provide recommendations on academic success strategies (e.g., tutoring, coaching, student learning preferences). All ARC services will continue to be provided online. Please visit the Office of Undergraduate Advising website for additional information related to academic advising and student support services.

Writing and Language Studio

For additional help with academic writing—and more generally with language learning—you are welcome to make an appointment with the Writing and Language Studio (WLS). To accommodate students who are learning remotely as well as those who are on campus, writing and language coaching appointments are available in person and online. You can register for an account, make an appointment, and learn more about WLS services, policies, and events on the WLS website.

IT Support

If you are experiencing technical difficulties, please contact IT:

- China-based faculty/staff/students 400-816-7100, (+86) 0512- 3665-7100
- US-based faculty/staff/students (+1) 919-660-1810
- International-based faculty/staff/students can use either telephone option (recommend using tools like Skype calling)
- Live Chat: https://oit.duke.edu/help
- Email: service-desk@dukekunshan.edu.cn

It is recommended that you familiarize yourself with DKU's VPN and proxy, since these may enable you to access resources you might be unable to reach otherwise. When using the VPN, you will usually want to select Duke VPN (portal.duke.edu), then INTL-DUKE.

What is the expected course schedule?

<u>Unit 1: Distributions</u>

(1)   2024-10-21   Class welcome

                       Readings

                       Chapter 1, 2.1 and 2.2, and 3.1 to 3.3

(2)   2024-10-23   Characteristics of distributions

                       Readings
                       Chapter 2.3-5 and 4

(3)   2024-10-28   Comparing distributions and the Normal distribution

                       Readings
                       Chapter 5

<u>Unit 2: Relationship between variables</u>

(4)   2024-10-30   Association and correlation

                       Readings

                       Chapter 6

(5)   2024-11-04   Simple Linear Regression

                       Readings
                       Chapter 7

(6)   2024-11-06   Regression Wisdom (outliers, lurking variables)

                       Readings
                       Chapter 8

(7)   2024-11-11   Multiple Regression

                       Readings
                       Chapter 9

        2024-11-13   Midterm exam

<u>Unit 3: Measuring uncertainty</u>

(8)   2024-11-18   Confidence intervals – proportions

                       Readings

Chapter 13

(9) 2024-11-20 Confidence intervals – means

Readings
Chapter 14

(10) 2024-11-25 Hypothesis testing

Readings
Chapter 15

(11) 2024-11-27 Hypothesis testing wisdom (p-values)

Readings
Chapter 16

Unit 4 – Statistical Inference

(12) 2024-12-02 Comparing groups

Readings
Chapter 17

(13) 2024-12-04 Returning to regression (regression assumptions)

Readings
Chapter 20

Lab schedule

(1) 2024-10-22 R and Quarto familiarization

(2) 2024-10-29 Advanced Quarto editing

(3) 2024-11-05 Working with regressions

(4) 2024-11-12 Interpreting coefficients

(5) 2024-11-19 Sampling

(6) 2024-11-26 Bootstrapping

–   2024-12-03 – No lab –

Assignment schedule

(1)	2024-10-27	DataCamp 1 (Introduction to R)
(2)	2024-11-03	DataCamp 2 (Introduction to the Tidyverse)
(3)	2024-11-10	DataCamp 3 (Introduction to Data Visualization with ggplot2)
(4)	2024-11-13	Midterm exam
(5)	2024-11-17	DataCamp 4 (Intermediate Data Visualization with ggplot2)
(6)	2024-11-24	DataCamp 5 (Modeling with Data in the Tidyverse)
(7)	2024-12-01	DataCamp 6 (Hypothesis Testing in R)
(9)	2024-12-09	Final project
	Lectures 2-13	Quiz